



پایان‌نامه‌ی کارشناسی ارشد: جواد ارجمند، ۱۳۹۸

## هوشمندسازی الگوریتم TF-IDF با استفاده از شبکه عصبی برای استخراج کلمات کلیدی از متون فارسی و تعیین موضوع بر اساس آن

با گسترش روزافزون حجم اطلاعات متنی، استخراج اطلاعات از این متن‌ها و پی بردن به موضوع و زمینه‌ی آن‌ها ضروری است. با توجه به اینکه ممکن است متن‌ها طولانی باشند پی بردن به اهداف آن‌ها دشوار است. بنابراین می‌توان از کلمات کلیدی که لغات مهم متن را فراهم می‌کنند برای استخراج و دست‌یابی به مفاهیم اصلی استفاده کرد. استخراج کلمات کلیدی از همه‌ی متن‌ها به صورت دستی دشوار است به همین دلیل به دنبال روشی هستیم تا بتوان کلمات کلیدی را به صورت هوشمند استخراج کرد. روش‌های متفاوتی برای استخراج کلمات کلیدی وجود دارند که هر یک دارای مزایا و معایبی هستند. در این پایان‌نامه از تکنیک‌های هوش مصنوعی مانند شبکه عصبی برای استخراج کلمات کلیدی استفاده شده است. با توجه به اینکه داده‌هایی که برای آموزش شبکه عصبی لازم هستند باید به صورت عددی باشند بنابراین لازم است قبل از استفاده از آن، پیش‌پردازشی بر روی متن انجام شود. در این پیش‌پردازش کلمات عمومی، کلمات کمتر از سه حرف و نشانه‌گذاری‌های متن حذف شده و یکسان‌سازی بر روی متن انجام می‌شود. حال می‌توان ورودی‌های شبکه عصبی را ایجاد کرد. در این پایان‌نامه از شبکه عصبی پرسپترون استفاده شده است که ورودی‌های این شبکه عبارت‌اند از تعداد تکرار هر کلمه در یک متن، تعداد تکرار هر کلمه در کل متن‌های پایگاه داده، تعداد تکرار هر کلمه در گروه متن‌های با موضوع یکسان، وجود کلمه در عنوان، وجود کلمه در پاراگراف اول متن و تعیین موضوع متن. با این ورودی‌ها و چند متن که از قبل کلمات کلیدی آن‌ها مشخص است، شبکه عصبی آموزش دیده و حال می‌توان از آن برای استخراج کلمات کلیدی همه‌ی متن‌ها استفاده کرد. نتایج نشان می‌دهد که روش پیشنهادی نتایج قابل قبول و با دقت بالایی را ایجاد کرده است.

**کلیدواژه‌ها:** استخراج اطلاعات، کلمه کلیدی، پیش‌پردازش متن، شبکه عصبی

شماره‌ی پایان‌نامه: ۱۲۷۴۱۰۱۵۹۶۲۰۰۶

تاریخ دفاع: ۱۳۹۸/۱۰/۳۰

رشته‌ی تحصیلی: مهندسی کامپیوتر- نرم‌افزار

دانشکده: فنی و مهندسی

استاد راهنما: مهندس کبری بهروان

استاد مشاور: مهندس علی اکبر نقابی

**M.A. Thesis:**

Intelligent TF-IDF algorithm Using Neural Network To



## Extract keywords from Persian texts And Determine the topic based on it

By increasing volume of textual information, it is necessary to extract information from these texts and to understand their subject and context. Given that texts may be long, it is difficult to understand their purpose. So keywords that provide important text vocabulary can be used to extract and access key concepts. It is difficult to extract keywords from all texts manually, so we are looking for a way to extract keywords intelligently. There are different ways to extract keywords that each have advantages and disadvantages.

This thesis uses artificial intelligence techniques such as neural network to extract keywords. Since the data needed for neural network training must be numerical, it is necessary to pre-process the text before using it. In this preposition of general words, words less than three letters and text markups are deleted and the text is merged. Now you can create neural network inputs.

This thesis uses Perceptron neural network whose inputs are number of repetitions of each word in a text, number of repetitions of each word in all database texts, number of repetitions of each word in the text group with The same theme is the text in the title, the text in the first paragraph of the text, and the topic of the text. With these inputs and some keywords already known, the neural network is trained and can now be used to extract keywords for all texts. The results show that the proposed method has higher accuracy than similar methods.